

Predictors of Earnings Risk with Machine Learning

Ethan Ballou* Scott Drewianka*

December 17, 2025

Abstract

This paper looks at the determinants of lifetime earnings risk under a Restricted Income Profile (RIP) model using traditional and machine learning methods such as lasso and SHAP values. The paper builds on the work of Drewianka and Oberg (2025) which uses a moment condition approach derive a parameter that captures permanent income risk. The paper finds that education and age are important in explaining lifetime earnings risk. The paper also finds that macroeconomic variables such as probability of recession and real GDP growth are important and along with state controls may further imply a role of government policy. Finally, the paper finds that occupation controls are important while industry controls do not appear to play a strong role.

Keywords: machine learning, restricted income profile, earnings instability, risk

JEL Codes: D8, J0, D3

*University of Wisconsin - Milwaukee

1. NEED TO ADD INFO ABOUT WHICH DUMMIES WERE INCLUDED IN STEPWISE REGRESSIONS SINCE THEY ARE SUPPRESSED ON THE TABLE
2. NEED TO ADD IN AND OCD Lasso, and of all of lasso for that matter
3. NEED TO ADD IN SHAP VALUE FIGURES
4. ONLY HAVE 25k observations and Scott had 62900

1 Introduction

2 Data and Model

3 Empirical Strategy

4 Results

5 Conclusion

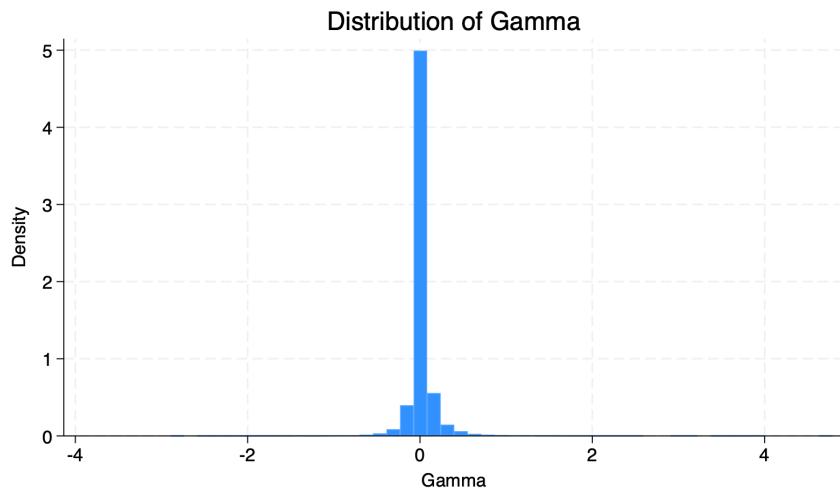


Figure 1: Distribution of Gamma

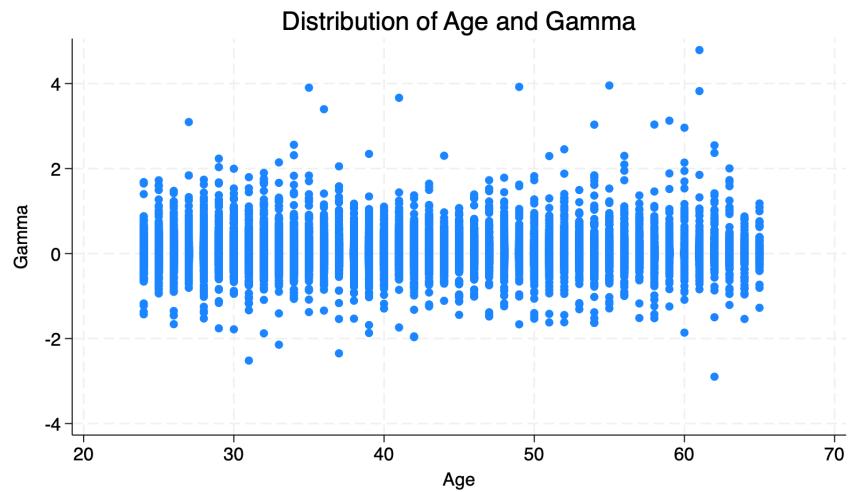


Figure 2: Scatterplot of Age vs. Gamma

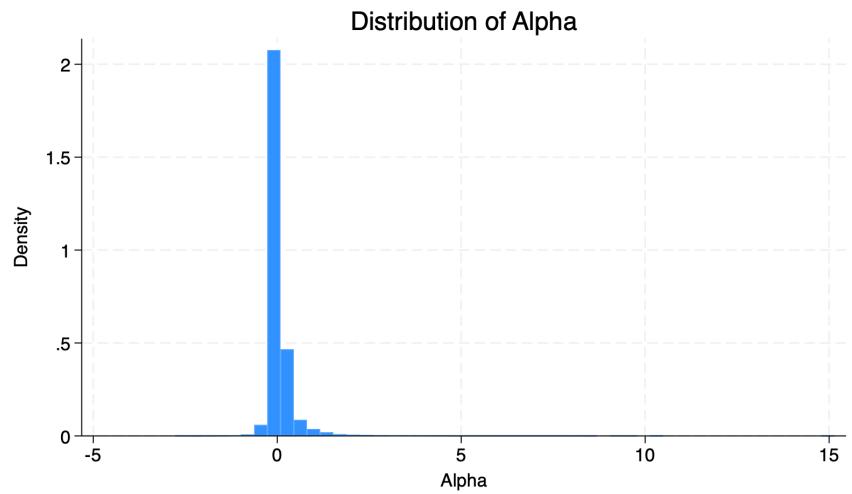


Figure 3: Distribution of Alpha

Table 1: Gamma Regressions: OLS Results

| | (1) Gamma | (2) Gamma | (3) Gamma | (4) Gamma | (5) Gamma |
|------------------------------|-------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|
| Less than High School | -0.0115*** (0.00215) | -0.0107*** (0.00232) | -0.0117*** (0.00242) | -0.0109*** (0.00253) | -0.0117*** (0.00255) |
| High School Graduate | -0.00918*** (0.00144) | -0.00854*** (0.00149) | -0.00902*** (0.00158) | -0.00812*** (0.00171) | -0.00850*** (0.00173) |
| Some College | -0.00746*** (0.00170) | -0.00681*** (0.00172) | -0.00736*** (0.00177) | -0.00674*** (0.00184) | -0.00695*** (0.00185) |
| Probability of Recession | 0.0000422 (0.0000437) | -0.0104 (0.0179) | -0.0000953 (51.52) | -0.0000817 (51.50) | -0.0000292 (51.49) |
| Real GDP growth rate | 0.000600 (0.000400) | -0.00432 (0.00325) | -0.0000211 (3.981) | -0.000186 (3.979) | -0.000626 (3.979) |
| 5-year moving average of AEP | 0.0000636* (0.0000295) | 0.0000629 (0.0000322) | 0.0000374 (0.0000340) | 0.0000766* (0.0000347) | 0.0000688 (0.0000356) |
| Out of Labor Force | -0.00430 (0.00535) | -0.00410 (0.00536) | -0.00471 (0.00573) | -0.00461 (0.00572) | -0.00474 (0.00572) |
| Tenure | -0.0000438 (0.0000888) | -0.0000336 (0.0000963) | 0.0000277 (0.0000981) | 0.00000375 (0.0000984) | 0.0000156 (0.0000987) |
| Age | 0.00592* (0.00269) | 0.00650* (0.00273) | 0.00679* (0.00274) | 0.00679* (0.00274) | 0.00671* (0.00274) |
| Age Squared | -0.000170** (0.0000651) | -0.000193** (0.0000659) | -0.000196** (0.0000662) | -0.000197** (0.0000662) | -0.000194** (0.0000663) |
| Age Cubed | 0.00000153** (0.000000509) | 0.00000172*** (0.000000515) | 0.00000173*** (0.000000517) | 0.00000173*** (0.000000518) | 0.00000171*** (0.000000518) |
| State FE | No | Yes | Yes | Yes | Yes |
| Year FE | No | Yes | Yes | Yes | Yes |
| Race FE | No | Yes | Yes | Yes | Yes |
| Cohort FE | No | Yes | Yes | Yes | Yes |
| Occupation FE | No | No | No | Yes | Yes |
| Industry FE | No | No | Yes | No | Yes |
| R-squared | 0.001 | 0.002 | 0.003 | 0.005 | 0.006 |
| N | 82357 | 82333 | 81556 | 81556 | 81556 |

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2: Gamma Regressions: Stepwise Selection Results

| | (1) Gamma | (2) Gamma | (3) Gamma | (4) Gamma | (5) Gamma |
|------------------------------|-------------------------------|-------------------------------|-------------------------------|-------------------------------|-------------------------------|
| Less than High School | -0.0110*** (0.00211) | -0.0110*** (0.00211) | -0.0101*** (0.00205) | -0.0108*** (0.00234) | -0.0110*** (0.00236) |
| High School Graduate | -0.00904*** (0.00143) | -0.00905*** (0.00143) | -0.00840*** (0.00143) | -0.00829*** (0.00166) | -0.00834*** (0.00168) |
| Some College | -0.00740*** (0.00169) | -0.00740*** (0.00169) | -0.00731*** (0.00172) | -0.00723*** (0.00182) | -0.00725*** (0.00183) |
| Age | 0.00571* (0.00268) | 0.00568* (0.00268) | | 0.00594* (0.00270) | 0.00566* (0.00270) |
| Age Cubed | 0.00000150** (0.000000508) | 0.00000149** (0.000000508) | 0.000000439*** (8.71e-08) | 0.00000150** (0.000000511) | 0.00000143** (0.000000511) |
| 5-year moving average of AEP | 0.0000579* (0.0000286) | 0.0000577* (0.0000286) | | 0.0000788* (0.0000316) | 0.0000732* (0.0000324) |
| Age Squared | -0.000165* (0.0000650) | -0.000165* (0.0000650) | -0.0000282*** (0.00000578) | -0.000168* (0.0000654) | -0.000161* (0.0000654) |
| Constant | -0.0391 (0.0358) | -0.0387 (0.0358) | 0.0395*** (0.00358) | -0.0414 (0.0360) | -0.0380 (0.0360) |
| State FE | | | | | |
| Year FE | | | | | |
| Race FE | | | | | |
| Cohort FE | | | | | |
| Occupation FE | | | | ✓ | ✓ |
| Industry FE | | | ✓ | | ✓ |
| State FE Available | No | Yes | Yes | Yes | Yes |
| Year FE Available | No | Yes | Yes | Yes | Yes |
| Race FE Available | No | Yes | Yes | Yes | Yes |
| Cohort FE Available | No | Yes | Yes | Yes | Yes |
| Occupation FE Available | No | No | No | Yes | Yes |
| Industry FE Available | No | No | Yes | No | Yes |
| R-squared | 0.001 | 0.001 | 0.002 | 0.003 | 0.004 |
| N | 82357 | 82333 | 81556 | 81556 | 81556 |

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3: Gamma Regressions: Lasso Selection Results

| | No Controls | No Occ/Ind | No Occ | No Ind | All Controls |
|----------------|-------------|------------|--------|--------|--------------|
| EDU1 | 2 | 3 | 3 | 2 | 2 |
| EDU2 | 1 | 2 | 2 | 2 | 2 |
| EDU3 | 3 | 4 | 4 | 4 | 3 |
| OLF | 9 | 9 | 8 | 7 | 6 |
| PrRecess | 10 | | | | |
| currentage | 8 | 7 | 6 | 5 | 4 |
| currentagecube | 4 | 8 | 7 | 6 | 5 |
| currentagesq | 11 | 10 | 10 | 9 | 8 |
| ma5aep | 5 | 5 | 5 | 3 | 3 |
| rGDPgrow | 6 | | | | |
| tenure | 7 | 6 | 9 | 8 | 7 |
| State FE | No | Yes | Yes | Yes | Yes |
| Year FE | No | Yes | Yes | Yes | Yes |
| Race FE | No | Yes | Yes | Yes | Yes |
| Cohort FE | No | Yes | Yes | Yes | Yes |
| Occupation FE | No | No | No | Yes | Yes |
| Industry FE | No | No | Yes | No | Yes |

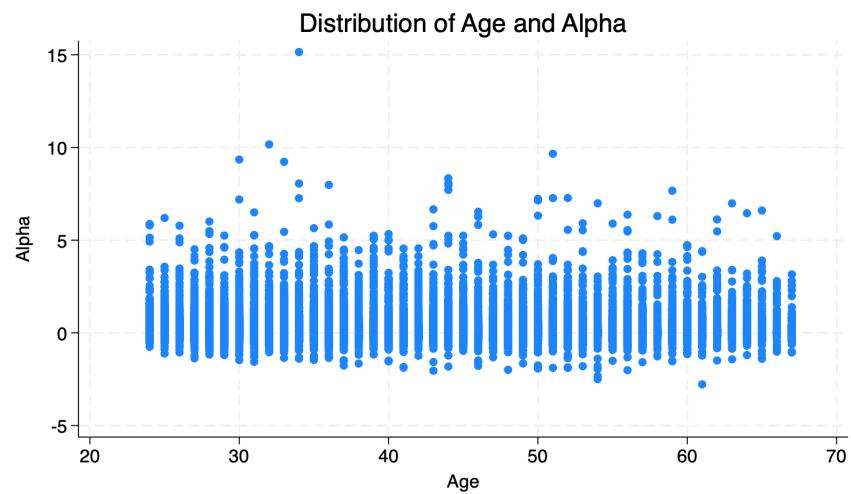


Figure 4: Scatterplot of Age vs. Alpha

Table 4: Alpha Regressions: OLS Results

| | (1) Alpha | (2) Alpha | (3) Alpha | (4) Alpha | (5) Alpha |
|------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|
| Less than High School | -0.0268*** (0.00451) | -0.0240*** (0.00487) | -0.0308*** (0.00508) | -0.0278*** (0.00531) | -0.0301*** (0.00535) |
| High School Graduate | -0.0131*** (0.00309) | -0.0129*** (0.00320) | -0.0136*** (0.00340) | -0.00909* (0.00368) | -0.0104** (0.00372) |
| Some College | -0.00240 (0.00366) | -0.00321 (0.00370) | -0.00223 (0.00384) | 0.00135 (0.00398) | 0.000562 (0.00401) |
| Probability of Recession | -0.0000420 (0.0000929) | -0.138* (0.0554) | -0.113 (0.0723) | -0.112 (0.0723) | -0.113 (0.0722) |
| Real GDP growth rate | -0.00144 (0.000855) | 0.000541 (0.00633) | -0.00107 (0.00713) | -0.00287 (0.00710) | -0.00359 (0.00714) |
| 5-year moving average of AEP | 0.00127*** (0.0000616) | 0.00138*** (0.0000669) | 0.00103*** (0.0000712) | 0.00111*** (0.0000726) | 0.00106*** (0.0000745) |
| Out of Labor Force | 0.113*** (0.0101) | 0.111*** (0.0101) | 0.0518*** (0.0111) | 0.0521*** (0.0111) | 0.0517*** (0.0111) |
| Tenure | -0.000698*** (0.000185) | -0.00000130 (0.000202) | 0.000135 (0.000207) | 0.0000259 (0.000208) | 0.0000861 (0.000208) |
| Age | 0.0165** (0.00520) | 0.0183*** (0.00526) | 0.0175*** (0.00530) | 0.0182*** (0.00530) | 0.0173** (0.00530) |
| Age Squared | -0.000398** (0.000124) | -0.000451*** (0.000125) | -0.000431*** (0.000126) | -0.000452*** (0.000126) | -0.000427*** (0.000126) |
| Age Cubed | 0.00000324*** (0.000000947) | 0.00000371*** (0.000000955) | 0.00000356*** (0.000000963) | 0.00000374*** (0.000000963) | 0.00000353*** (0.000000964) |
| State FE | No | Yes | Yes | Yes | Yes |
| Year FE | No | Yes | Yes | Yes | Yes |
| Race FE | No | Yes | Yes | Yes | Yes |
| Cohort FE | No | Yes | Yes | Yes | Yes |
| Occupation FE | No | No | No | Yes | Yes |
| Industry FE | No | No | Yes | No | Yes |
| R-squared | 0.008 | 0.013 | 0.020 | 0.021 | 0.023 |
| N | 102946 | 102910 | 100781 | 100781 | 100781 |

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: Alpha Regressions: Stepwise Selection Results

| | (1) Alpha | (2) Alpha | (3) Alpha | (4) Alpha | (5) Alpha |
|------------------------------|--------------------------------|--------------------------------|-------------------------------|--------------------------------|-------------------------------|
| Less than High School | -0.0261*** (0.00418) | -0.0204*** (0.00450) | -0.0259*** (0.00464) | -0.0259*** (0.00477) | -0.0272*** (0.00480) |
| High School Graduate | -0.0122*** (0.00271) | -0.0108*** (0.00278) | -0.0107*** (0.00289) | -0.00852** (0.00304) | -0.00913** (0.00306) |
| Age Squared | -0.000401** (0.000124) | -0.000394** (0.000124) | -0.000336** (0.000125) | -0.000384** (0.000125) | -0.000351** (0.000125) |
| Age Cubed | 0.00000326*** (0.000000947) | 0.00000327*** (0.000000950) | 0.00000287** (0.000000959) | 0.00000321*** (0.000000959) | 0.00000296** (0.000000959) |
| Real GDP growth rate | -0.00115* (0.000541) | | | | |
| 5-year moving average of AEP | 0.00127*** (0.0000601) | 0.00135*** (0.0000650) | 0.00106*** (0.0000689) | 0.00111*** (0.0000711) | 0.00107*** (0.0000728) |
| Out of Labor Force | 0.113*** (0.0101) | 0.112*** (0.0101) | 0.0535*** (0.0111) | 0.0541*** (0.0111) | 0.0535*** (0.0111) |
| Tenure | -0.000704*** (0.000185) | | | | |
| Age | 0.0167** (0.00519) | 0.0159** (0.00523) | 0.0132* (0.00527) | 0.0153** (0.00527) | 0.0140** (0.00527) |
| Constant | -0.190** (0.0704) | -0.157 (0.0825) | -0.102 (0.0831) | -0.117 (0.0831) | -0.105 (0.0831) |
| State FE | | ✓ | ✓ | ✓ | ✓ |
| Year FE | | ✓ | ✓ | ✓ | ✓ |
| Race FE | | ✓ | ✓ | ✓ | ✓ |
| Cohort FE | | | | | |
| Occupation FE | | | | ✓ | ✓ |
| Industry FE | | | ✓ | | ✓ |
| State FE Available | No | Yes | Yes | Yes | Yes |
| Year FE Available | No | Yes | Yes | Yes | Yes |
| Race FE Available | No | Yes | Yes | Yes | Yes |
| Cohort FE Available | No | Yes | Yes | Yes | Yes |
| Occupation FE Available | No | No | No | Yes | Yes |
| Industry FE Available | No | No | Yes | No | Yes |
| R-squared | 0.008 | 0.012 | 0.018 | 0.021 | 0.022 |
| N | 102946 | 102910 | 100781 | 100781 | 100781 |

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$